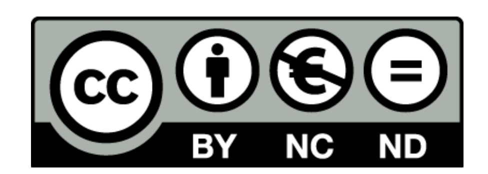
Métodos de análisis de datos
Problemas

Zenaida Hernández Martín
Esther Raya Díez

Esta publicación ha sido financiada por la Universidad de La Rioja, dentro del proyecto de innovación docente APIDUR nº 10 – 2013/14, “Formación en competencias para la práctica profesional en el Grado en Trabajo Social”.



ISBN 978-84-606-9653-7
Edita: Universidad de La Rioja

Prólogo

El plan de estudios de Grado en Trabajo Social de la Universidad de la Rioja presenta tres elementos característicos que lo diferencian de la oferta de otras universidades: por un lado proporciona una amplia formación jurídica, por otro lado ofrece una formación básica en Ciencias Jurídicas y Sociales común a otros grados (Derecho, GADE y Relaciones Laborales y Recursos Humanos) y en tercer lugar desarrolla las prácticas externas basadas en la metodología de Aprendizaje Servicio.

Estos tres elementos pretenden ser una fortaleza, en cuanto a las competencias desarrolladas por los estudiantes, para su futura inserción laboral en un entorno socioeconómico complejo, que demanda profesionales polivalentes.

En determinadas materias, como la Estadística, es importante que los estudiantes perciban la relación del contenido con las competencias específicas del Grado. En este sentido, durante el cuso 2014-2015 se desarrolló el proyecto de innovación docente (correspondiente a la convocatoria 2013-2014): "Formación en competencias para la práctica profesional en el Grado en Trabajo Social". Uno de los objetivos de dicho proyecto era "Elaborar materiales docentes orientados a la actividad a desarrollar en las prácticas externas, que mejoren el rendimiento académico en asignaturas básicas".

Este documento es el resultado de la actividad llevada a cabo por el equipo de elaboración de materiales docentes de dicho proyecto y se ha realizado mediante la adaptación de los problemas que ya se resolvían en la asignatura de Métodos de análisis de datos, para darles un perfil más acorde con la titulación de Trabajo Social. Todo ello con el objetivo de acercar más esta materia a los estudiantes.

En la bibliografía de referencias sobre Estadística se constata una ausencia de materiales especializados en Trabajo Social, con ejemplos aplicados al perfil profesional de la titulación. Nuestro próximo objetivo es ir, poco a poco, mejorando esta situación.

Este libro de problemas tiene por finalidad comenzar a suplir esta carencia y ayudar a los estudiantes en su proceso de aprendizaje en el marco de su titulación.

Logroño, diciembre de 2015

Índice general

Problemas de: Estadística descriptiva unidimensional

1. Identifica, de las siguientes variables, cuáles son continuas y cuáles discretas:
   - Hora del día.
   - Tiempo de espera en una sala de urgencias.
   - Número de horas de servicio de ayuda a domicilio.
   - Peso del alimento ingerido.
   - Número usuarios de un centro de día.
   - Edad (años) de los sujetos de un experimento.
   - Velocidad de un corredor en una carrera.

2. Identifica la escala de las siguientes variables:
   - Fecha de nacimiento.
   - Tipos de bicicletas utilizadas por los estudiantes de primer curso.
   - Tiempo que se tarda en reaccionar a un sonido.
   - Peso de un grupo de mujeres.
   - Tipo de prestaciones de un servicio social comunitario
   - Eficiencia en Estadística, clasificada como: mala, regular y buena.
   - Calificación de los estudiantes del curso en el examen de Estadística.

3. Construye la tabla de frecuencias (absoluta, relativa y acumuladas), para la siguiente masa de datos:

   25, 22, 21, 22, 26, 29, 29, 27, 25, 21, 22, 25, 26, 24, 24, 25, 26, 29, 29, 21, 26, 25, 23, 29, 28, 26, 24, 21, 23, 25, 21, 21, 25, 27, 27, 28, 26, 27, 22, 23, 25, 22, 26, 28, 27, 28, 24, 26, 27, 29, 21, 22, 21, 25, 24, 29, 23, 25, 22, 23.

4. Convierte la siguiente distribución, en una distribución de frecuencias con 4 intervalos, y

   a) Determina el tamaño de los intervalos.

   b) Indica los límites inferior y superior de cada intervalo.

   c) Identifica la marca de clase de cada intervalo.

d) Encuentra la frecuencia de cada intervalo.

e) Calcula la frecuencia relativa acumulada para cada intervalo.

x_i	1	2	3	4	5	6	7	8	9	10	11	12	
n_i	2	1	1	2	3	4	5	6	5	4	4	3	40

5. De un examen realizado a un grupo de alumnos, cuyas notas se han evaluado del 1 al 8, se ha obtenido el siguiente cuadro estadístico:

x_i	n_i	N_i	f_i	F_i
1	4		0.08	
2	4			
3		16	0.16	
4	7		0.14	
5	5	28		
6		38		
7	7	45	0.14	
8				

Se pide:

a) Termina de rellenar la tabla

b) Nº de alumnos que se han examinado

c) Nº de alumnos que han obtenido una nota superior a 3

d) % de alumnos que han sacado una nota igual a 6

e) Nº de alumnos que han obtenido una nota superior a 2 e inferior a 5.

6. La siguiente tabla muestra los resultados de una encuesta sobre hábitos saludables:

	Fumador	No fumador
Hombres	93	46
Mujeres	90	127
Total	183	173

Determina:

a) Porcentaje de mujeres fumadoras.

b) Proporción de hombres.

c) De los fumadores del estudio, la proporción de hombres.

d) Porcentaje de fumadoras entre las mujeres.

7. En un centro de día se encuentran 4 usuarios con nivel de dependencia severa y 24 usuarios con nivel de dependencia moderada, ¿cuál es la razón de usuarios con nivel de dependencia severa y moderada?

8. En un grupo de 125 hombres y 80 mujeres, ¿cuál es la razón de hombres a mujeres?

9. Si ocurren 200 nacimientos entre 5400 mujeres en edad de concebir, ¿cuál es la tasa de nacimientos?

10. ¿Cuál es la tasa de cambio para un aumento de la población de 1500 en 1985 a 2700 en el año 2005?

11. Analiza el siguiente comentario: "La mejor medida de tendencia central es la media aritmética, por eso la utilizaremos siempre salvo que no se conozcan los valores extremos de la variable".

12. Con motivo de la venta de una casa, un agente inmobiliario estableció que el ingreso medio en la zona era de 12.000 euros, sin embargo, en una reunión de protesta de contribuyentes de la zona (donde él también reside) afirmó que el ingreso más común era de 7.200 euros ¿Puede explicarse esta aparente contradicción?, ¿es necesariamente deshonesto ese agente?.

13. Suponemos que el ingreso anual disponible de los 5 millones de habitantes de cierto país tienen una media de 4.800 u.m. y una mediana de 3.400 u.m.

    a) ¿Cuál es el ingreso nacional disponible total anual?
    b) ¿Cuál considerarías que es el ingreso anual disponible del *habitante tipo*?

14. Los salarios por hora de los 7 empleados de una pequeña empresa de ayuda a domicilio son: 12, 15, 18, 13, 13, 10 y 16 euros respectivamente. Responde a las siguientes cuestiones:

    a) ¿Cuál es el salario por hora más habitual?
    b) ¿Qué salario no es superado por la mitad de los empleados?
    c) ¿Cuál es el salario medio por hora, en esta empresa?

15. La distribución de las acciones de una determinada sociedad viene dada de la siguiente forma:

Número de acciones	Número de accionistas
0 - 20	10
20 - 28	32
28 - 32	50
32 - 48	8

    Se pide:

    a) El capital aproximado de la empresa, supuesto un valor nominal para cada acción de 500 euros.
    b) Nº medio de acciones por accionista.
    c) Paquete de acciones poseídas por un mayor número de accionistas.
    d) En el supuesto de que en la junta general de accionistas los votos se establecen en proporción al número de acciones poseídas, ¿qué mínimo de acciones debe tener un accionista para que su poder decisorio sea mayor al de la mitad de los socios?

16. Dados los valores observados: 20, 22, 25, 29, 33, 34, 38, 40; determina, para cada uno de ellos cuál es su desviación respecto a la media del conjunto. ¿Cuánto suman las desviaciones?

17. Se considera que, a nivel universitario, la homogeneidad en la edad de los estudiantes de un grupo es un factor que influye de forma positiva en su aprendizaje. Si las edades de los estudiantes de dos grupos son las que se dan a continuación:

Mañana	23	29	27	22	24	21	25	26	27	24
Tarde	27	34	30	27	28	30	34	35	28	29

¿En cuál de los dos grupos será más fácil enseñar?

18. Los salarios mensuales de los auxiliares de dos residencias de personas mayores, A y B, son los que se dan en la siguiente tabla:

Salarios (€)	Residencia A	Residencia B
$[550, 750)$	10	7
$[750, 1050)$	32	20
$[1050, 1550)$	57	37
$[1550, 2550)$	54	78

Determina:

a) El salario medio en cada una de las dos residencias.

b) El salario más frecuente en cada una de las dos residencias.

c) ¿Cuál es el salario que no es superado por el 50 % en cada una de las dos residencias?

d) ¿En cuál de las dos residencias hay mayor homogeneidad salarial?, ¿por qué?

19. Razona la verdad o falsedad de las siguientes afirmaciones:

a) El intervalo modal de una distribución agrupada en intervalos es aquel que tiene mayor frecuencia absoluta.

b) Si la los precios de los servicios sanitarios aumentan en un 10 %, la varianza de esos precios aumentará en un 21 %.

20. Los salarios de los 50 trabajadores de una residencia sanitaria se recogen en la siguiente tabla:

Salario (u.m.)	85	100	110	115	150	185	250
Nº trabajadores	6	7	12	11	8	4	2

La masa salarial total, en la residencia, ¿está repartida de forma equitativa?, ¿por qué?

21. Se dispone de la siguiente información respecto a las ganancias medias mensuales de los trabajadores en residencias de ancianos en un determinado país:

Ganancias (miles de u.m.)	25-30	30-40	40-50	50-60	60-80	80-100
Trabajadores (en miles)	34	211	332	422	472	194

a) Dibuja la curva de Lorenz e indica si hay concentración salarial entre los trabajadores de las residencias de ancianos.

b) Calcula el índice de concentración de Gini e interprétalo.

Problemas de: Números índices

1. En un residencia, las cifras de ingresos anuales por las cuotas de los usuarios en el periodo 2008/2012, en miles de euros, fueron:

Años	Ingresos
2008	118
2009	124
2010	128
2011	136
2012	146

Obtén:

a) La serie de índices que nos dan la relación de los ingresos por cuotas de cada año respecto al año 2008.

b) La serie de índices que nos dan la relación de los ingresos por cuotas de cada año respecto al año anterior.

c) La serie de índices que nos dan la relación de los ingresos por cuotas de cada año respecto al año 2010

2. En una empresa de prestación de servicios asistenciales, las cifras de ventas anuales, por el servicio de comida a domicilio, para el periodo 2010/2014, expresadas en porcentajes, en relación con las del año anterior fueron:

Años	I_{t-1}^{t}
2010	118
2011	104
2012	98
2013	96
2014	106

Sabiendo que en el año 2009, las ventas alcanzaron los 10 millones de euros, determina:

a) Las cifras de ventas de cada uno de los años.

b) El índice de ventas, con base en el año 2009.

3. En 2012 la demanda del servicio de ayuda a domicilio creció un 40 % sobre la de 2011. En 2013 la demanda de este servicio estaba un 20 % por debajo de la de 2012, pero un 16.67 % por encima de la de 2014. Determina las relaciones de demanda para los años, desde 2011 a 2014, tomando como base 2014.

4. Una magnitud aumenta, entre el año 0 y el año 1, en un 30 %, entre el año 1 y el año 2 en un 20 % y ha disminuido entre el año 2 y el año 3 en un 40 %. Obtener la serie de números índices para esta magnitud con base el año 0.

5. Los precios de ciertos productos lácteos y las cantidades consumidas por una familia de bajos recursos en los años 2012 y 2013 son los siguientes:

	Precios en euros		Cantidades	
Artículos	2012	2013	2012	2013
Leche	0, 7 /l	0,75 /l	1.000 l	1.000 l
Queso	15 /Kg	20 /Kg	30 Kg	40 Kg
Mantequilla	8 /Kg	8,5 /Kg	40 Kg	50 Kg

Calcula los índices de precios y cantidades de Laspeyres, Paasche, y Fisher con base en el año 2012.

6. Los datos de que disponemos sobre el IPC de un determinado país, son los siguientes: en 2010 era 120; en 2011, 132; en 2012 el incremento anual fue del 9 %; en 2013 fue del 7 % y en 2014 fue del 6 %.

   Suponiendo que el índice de precios cumple todas las buenas propiedades de los índices, responde a lo siguiente:

   a) Obtén la serie completa con base en 2010.

   b) Si un alquiler se pactó en 600 u.m. en el año 2011 con una cláusula de revisión según el IPC, ¿cuánto se pagó en 2012?

   c) El alquiler pactado en 2011, a qué cantidad equivaldría en el año 2014?

   d) Si para los artículos de la cesta de la compra, el índice de valor para el año 2014 con base en el año 2012 es 115, ¿Cuál será el índice de cantidad en el mismo período?

7. La siguiente tabla muestra los índices de precios en un determinado país

Año	2000	2001	2002	2003	2004	2005	2006	2007
IPC	134.7	160.1	174.9	183.0	194.2	209.3	235.6	268.8

   Suponiendo que el índice de precios cumple todas las buenas propiedades de los índices, responde a lo siguiente:

   a) Obtén la serie de índices de precios con base en el año 2000.

   b) Obtén la serie de índices de precios tomando como base (período de referencia) la media aritmética de los índices de los años 2003 a 2005.

   c) Determina el poder adquisitivo de una unidad monetaria cada uno de los años de este período (toma como referencia el año 2000).

   d) Si en 2000 se hace una inversión de 100 u.m., en 2003 otra de 150 u.m. y en 2005 otra de 200 u.m., ¿a cuánto equivale la inversión en términos monetarios de 2007?

8. En la tabla siguiente se dan los salarios anuales, en euros, de los auxiliares de un centro de día, así como los índices de precios al consumo en el mismo periodo.

Años	IPC_{00}	Salarios
2002	170	20350
2003	173	21250
2004	178	22000
2005	186	23150
2006	189	24000
2007	196	25610

Suponiendo que el índice de precios cumple todas las buenas propiedades de los índices, responde a lo siguiente:

a) Calcula los índices, con base 2002, que miden la variación de los salarios.

b) Calcula los índices de precios de consumo con base 2002.

c) Calcula los salarios en u.m. constantes de 2002.

d) En términos constantes, ¿cuál ha sido la variación de los salarios de 2006 a 2007?

Problemas de: La curva Normal

1. Si la distribución de los salarios de los trabajadores sociales de una ONG, sigue una distribución Normal de media 1500 euros y desviación típica 350 euros:

   a) ¿En qué rango de valores se encuentran el 68.27 % central de las observaciones?

   b) ¿En qué rango de valores se encuentran el 95.45 % central de las observaciones?

   c) Teóricamente, ¿qué proporción de salarios se encuentran por debajo de los 800 euros?

   d) Teóricamente, ¿qué proporción de salarios se encuentran por encima de los 1850 euros?

2. Queremos comparar las calificaciones de un estudiante en dos asignaturas que tienen diferente escala. En la asignatura de Economía la nota media del grupo ha sido de 5.3, con una desviación típica de 2.2, mientras que en la asignatura de Informática la nota media ha sido de 66 con una desviación típica de 24 puntos.

   a) Si el estudiante ha sacado un 6.6 en Economía y un 78 en Informática, ¿en cuál de las dos asignaturas ha conseguido mejor calificación dentro del grupo?

   b) Si nos dicen que las calificaciones estandarizadas de un segundo estudiante han sido de un (-0.25) en Informática y de un 1.2 en Economía, ¿cuáles han sido sus calificaciones en estas asignaturas?.

   c) ¿En cuál de las dos asignaturas las calificaciones han sido más homogéneas?

3. Utilización de la tabla de la curva Normal tipificada (N(0,1)):

   Suponiendo que tenemos tipificados los gastos en servicios sociales de todos los Ayuntamientos de España, determina:

   a) Proporción de Ayuntamientos cuyo gasto es inferior a 0.45.

   b) Proporción de Ayuntamientos cuyo gasto es inferior a -0.53.

   c) Proporción de Ayuntamientos cuyo gasto es superior a 0.38.

   d) Proporción de Ayuntamientos cuyo gasto es superior a -0.45.

   e) Proporción de Ayuntamientos cuyo gasto se encuentra entre -0.36 y 0.45.

   f) Proporción de Ayuntamientos cuyo gasto se encuentra entre -0.36 y 0.36.

g) Proporción de Ayuntamientos cuyo gasto se encuentra entre 0.36 y 0.45.

h) Proporción de Ayuntamientos cuyo gasto se encuentra entre -0.36 y -0.15.

4. Si los valores de una variable se distribuyen según una Normal de media 7 y desviación típica 2, determina:

   a) Por debajo de qué valor se encuentran el 40 % de las observaciones.

   b) Por encima de qué valor se encuentran el 50 % de las observaciones.

   c) Entre qué valores se encuentra el 60 % central de las observaciones.

   d) Por debajo de qué valor se encuentran el 62 % de las observaciones.

   e) Por encima de qué valor se encuentran el 18 % de las observaciones.

   f) Entre qué valores se encuentra el 34 % central de las observaciones.

   g) ¿Qué porcentaje de observaciones se encuentran por debajo del 6? (Es decir, el valor 6, a qué percentil corresponde?).

   h) ¿Qué porcentaje de observaciones se encuentran por encima del 8?

5. Si la variable T se distribuye según un modelo t de Student con 20 grados de libertad, calcula el porcentaje de observaciones que:

   a) Son mayores que 1.064.

   b) Son menores que 0.3909.

   c) Se encuentran entre -1.7247 y 0.

6. Si la variable T se distribuye según una t de Student con 12 grados de libertad, determina qué valor o valores cumplen las siguientes condiciones (independientes):

   a) La proporción de valores mayores que él es 0.20.

   b) El porcentaje de valores menores que él es 2.5.

   c) El 80 % central de las observaciones está acotado por ellos.

Problemas de: Probabilidad y variables aleatorias

1. En un grupo de 1000 sujetos se pasó un test de inteligencia y se midió su rendimiento en una prueba, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia y según fuese apto o no apto en la prueba de rendimiento académico.

   Se encontraron 500 aptos, de los que 300 tenían inteligencia superior. De los 400 con inteligencia superior, 300 resultaron aptos.

   Extraemos un sujeto al azar y definimos los sucesos, A: *Ser superior a la media en inteligencia* y B: *Ser apto en rendimiento.* Determina:

   a) Si los sucesos A y B son independientes.

   b) Si seleccionamos un sujeto que resulta ser superior en inteligencia, ¿cuál es la probabilidad de que sea apto?

2. En una determinada población la probabilidad de que se den trastornos neuróticos (N) es 0.7; la probabilidad de que se den trastornos depresivos (D) es 0.2 y la probabilidad de que se den ambos a la vez es 0.1. Si tomamos un individuo al azar de esta población ¿cuál es la probabilidad de que tenga sólo uno de los dos trastornos?

3. Con el fin de estudiar la eficacia de tres tratamientos de las fobias, un experimentador divide a los enfermos en tres grupos: a cada grupo le va a aplicar uno de los tres tratamientos y la selección del mismo se llevará a cabo por procedimientos aleatorios.

   Lanza al aire una moneda y un dado. Si sale cara y un número par, se asigna el tratamiento A. Si sale cruz y número impar, se asigna el tratamiento B. Y en cualquier otro caso. el C.

   ¿Cuáles son las probabilidades de asignación de cada tratamiento con semejante procedimiento?

4. Tenemos una población en la que los grupos sanguíneos se distribuyen entre los sujetos de la siguiente forma: A: 45 %, B: 8 %, AB: 4 %, O: 43 %.

   Con respecto al factor RH, la quinta parte de los sujetos del grupo A tienen el factor RH negativo, así como la mitad de los del grupo B, la cuarta parte de los del grupo AB y la décima parte de los del grupo O. Responde a las siguientes cuestiones:

a) Si extraemos una observación al azar, ¿cuál ea la probabilidad de que sea del tipo A o tenga factor RH positivo?.
b) Si extraemos un sujeto al azar y resulta tener el factor RH positivo. ¿cuál es la probabilidad de que pertenezca al grupo O?

5. Un agente de seguros hace una póliza a un matrimonio en el que ambos cónyuges tienen 50 años. Según los estudios de esperanza de vida, la probabilidad de que un hombre de 50 años siga vivo dentro de 25 años es 0.40, mientras que la de que siga viva una mujer de esa edad es 0.50. Supuesto esto, y que la longevidad de los cónyuges es independiente la una de la otra, determina cuanto valen las probabilidades siguientes:
   a) La de que él esté vivo dentro de 25 años.
   b) La de que los dos estén vivos dentro de 25 años.
   c) La de que esté vivo al menos uno de ellos dentro de 25 años.

6. Un sujeto gana 200 euros si es capaz de detectar correctamente una señal acústica sobre un ruido de fondo y pierde 1.000 euros en cualquier otro caso (omisión, detección incorrecta. etc.). Sabiendo que la probabilidad de detección de la señal es 0.99, determina cuál es el beneficio medio por intento de detección en cada emisión de señal.

7. Para obtener fondos, con los que financiar el paso del ecuador, una promoción de estudiantes organiza una lotería con 1.000 papeletas que vende a 10 euros cada una.

   Si se entrega un primer premio de 2000 euros, dos segundos de 500 euros y diez reintegros:
   a) Determina el valor esperado de la variable *euros ganados* por papeleta.
   b) Manteniendo las demás cantidades constantes, ¿cuánto habría que subir el primer premio para que este juego fuese justo?

8. Organizamos un sorteo con un coche valorado en 6000 euros como primer premio, un televisor valorado en 600 euros como segundo premio y cinco bicicletas valoradas en 120 euros como terceros premios. Determina a cuanto tendríamos que vender cada uno de los 10.000 billetes para que este sea un juego justo.

9. En un servicio de ludoteca por horas, sabemos que en una determinada franja horaria, si llueve acuden 10 niños, pero si no llueve acude sólo uno. Si en esta época del año la probabilidad de que llueva es 0.25, determina cuántos niños esperaremos que acudan el próximo lunes en esa franja horaria.

10. Al estudiar las características físicas del personal que atiende a los ancianos en una residencia de personas mayores se ha observado que las alturas de estos trabajadores se distribuyen según una N(170,10) y su peso se distribuye según una N(75,5). Suponiendo que ambas variables son independientes en dicha población, si seleccionamos al azar un sujeto,

a) ¿Cuál es la probabilidad de que su altura sea menor que 168 cm?

b) ¿Cuál es la probabilidad de que su peso sea superior a 82 Kg?

c) ¿Cuál es la probabilidad de que su altura sea superior a 172 cm y su peso superior a 72 Kg?

11. Suponiendo que el cociente intelectual (CI) se distribuya según una N(100,15) en la población de estudiantes universitarios y elijamos aleatoria e independientemente una muestra de 1.000 sujetos:

    a) ¿**Cuántos** cabe esperar que se encuentren entre 85 y 115?

    b) ¿**Cuántos** por encima de 130?

12. En un examen valorado sobre 10 puntos las calificaciones de un grupo de 250 estudiantes se distribuyen aproximadamente según una normal, con media 5.6 y varianza 2.25.

    a) ¿Cuántos sujetos no pasan de 6?

    b) ¿Cuántos aprueban?

    c) Si se han dado 14 sobresalientes, determina qué puntuación mínima se ha exigido para darlo.

Problemas de: Inferencia estadística

1. Se sabe que la desviación típica de las notas de cierto examen de Matemáticas es 2,4. Para una muestra de 36 estudiantes se obtuvo una nota media de 5,6.

   a) Determina un intervalo de confianza para la nota media, con un nivel de confianza del 95 %.

   b) ¿Podemos confirmar la hipótesis de que la nota media del examen fue de 6, con un nivel de confianza del 95 %?

2. Un sociólogo ha pronosticado, que en una determinada ciudad, el nivel de abstención en las próximas elecciones será del 40 % como máximo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar.

   Determinar con un nivel de significación del 1 %, si se puede admitir el pronóstico.

3. Un informe indica que el precio medio del billete de avión entre Canarias y Madrid es, como máximo, de 120 euros con una desviación típica de 40 euros. Se toma una muestra de 100 viajeros y se obtiene que la media de los precios de sus billetes es de 128 euros.

   ¿Se puede aceptar, con un nivel de significación igual a 0,1 la afirmación de partida?

4. El director del servicio de teleasistencia afirma que, como máximo, el 6 % de las llamadas recibidas en el servicio son involuntarias, por parte de la persona mayor que lleva el dispositivo. Se revisaron 300 llamadas al azar y se detectaron 21 llamadas involuntarias.

   Con un nivel de significación del 1 %, ¿se puede aceptar la afirmación del director del servicio de teleasistencia?

5. La duración de la bombillas de 100 W que fabrica una empresa sigue una distribución Normal con una desviación típica de 120 horas de duración.

   Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. Con un nivel de significación de 0,01, ¿habría que rechazar el lote por no cumplir la garantía?

6. Un fabricante de lámparas eléctricas está ensayando un nuevo método de producción que se considerará aceptable si las lámparas obtenidas por este método dan

lugar a una población Normal de duración media 2400 horas, con una desviación típica igual a 300.

Se toma una muestra de 100 lámparas producidas por este método y esta muestra tiene una duración media de 2320 horas. ¿Se puede aceptar la hipótesis de validez del nuevo proceso de fabricación con un riesgo igual o menor al 5 %?

7. El control de calidad una fábrica de pilas y baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Hasta ahora el tiempo de duración en conversación seguía una distribución normal con media 300 minutos y desviación típica 30 minutos. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que de una muestra de 60 baterías el tiempo medio de duración en conversación fue de 290 minutos. Suponiendo que ese tiempo sigue siendo Normal con la misma desviación típica:

   ¿Se puede concluir que las sospechas del control de calidad son ciertas a un nivel de significación del 2 %?

8. En el programa de turismo social para mayores del Imserso están preocupados por la salubridad de las playas que se van a ofrecer este año. Para estar más tranquilos han encargado un estudio de salubridad y durante 10 días se han efectuado análisis del contenido de cierta sustancia tóxica en la playa de Gandía. Análogamente se ha procedido durante 8 días diferentes en la playa del Sardinero. En la siguiente tabla se muestran los niveles de toxicidad obtenidos cada día en ambas playas:

Playa	Niveles de sustancia tóxica									
Gandía	10	10	12	13	9	8	12	12	10	14
El Sardinero	11	8	9	7	10	8	8	10		

   a) Determina el intervalo de confianza al 90 % para el nivel de toxicidad diario medio en la playa de Gandía.

   b) Determina el intervalo de confianza al 90 % para el nivel de toxicidad diario medio en la playa de El Sardinero.

   c) c) Analiza, al 90 % de confianza, si el nivel tóxico medio es el mismo en ambas playas. ¿Es significativa la diferencia entre los niveles medios observados?

9. Una empresa suministradora del servicio de cocina a domicilio para personas mayores está considerando la decisión de adquirir nuevas máquinas etiquetadoras. Para comprobar si las nuevas máquinas mejoran significativamente la eficiencia de los empleados (colocan más etiquetas), selecciona aleatoriamente dos grupos de 9 trabajadores para realizar un control sobre el número de etiquetas colocadas en períodos de 5 minutos. En uno de los grupos, se utilizan las antiguas máquinas, y en el otro las nuevas. Tras un período de adaptación de los empleados, los resultados obtenidos son los siguientes:

Grupo 1 (antiguas)	305	312	300	284	290	264	272	301	275
Grupo 2 (nuevas)	303	301	310	303	309	296	315	282	272

Responde a los siguientes apartados, **enumerando y justificando en cada uno de ellos las condiciones de validez**:

a) Determina, a partir de la muestra, un intervalo de confianza al 90 % para el número medio de etiquetas con las máquinas antiguas.

b) Determina, a partir de la muestra, un intervalo de confianza al 95 % para el número medio de etiquetas con las máquinas nuevas.

c) ¿Podemos aceptar, para un nivel de significación del 5 %, que el número medio de etiquetas con las máquinas antiguas es de 295?

d) ¿Para qué nivel de significación podemos aceptar que el número medio de etiquetas colocadas por las máquinas nuevas es de 302?

e) ¿Podemos aceptar, con un nivel de significación del 1 % que el número medio de etiquetas colocadas por las dos máquinas es el mismo?

Problemas de: Tamaño de la muestra

1. El gerente de una residencia de personas mayores están interesado en conocer la temperatura media que alcanza un sistema de calefacción experimental. Suponiendo que dicha temperatura sigue una distribución Normal, con una desviación típica de 0.3 grados,

   a) determina cuál debe ser el tamaño de la muestra necesario, para construir un intervalo de confianza al 95 % con una longitud menor o igual que 0.1.

   b) determina cuál debe ser el tamaño de la muestra necesario, para construir un intervalo de confianza al 90 % con una longitud menor o igual que 0.15.

   c) determina cuál debe ser el tamaño de la muestra necesario, para construir un intervalo de confianza al 95 % con si queremos que el error muestral sea menor o igual que 0.1.

2. Una empresa de reparto de comida a domicilio asegura que, como máximo, el 35 % de sus pedidos sufren algún retraso. Para comprobarlo, vamos a contrastar esta hipótesis asumiendo un riesgo de equivocarnos del 5 %. Si estamos dispuestos a asumir un error muestral máximo de 0.05, ¿cuál debe ser el tamaño de nuestra muestra?

3. Se está realizando una investigación para determinar de qué manera afecta la administración de una hormona de crecimiento a la ganancia de peso de ratas preñadas. Se anotaron las ganancias de peso durante el período de gestación para un grupo de control de 6 ratas y para otro grupo de 6 ratas a las que se les administró la hormona del crecimiento, observándose unas ganancias medias de peso de 41.8 y 60.8 gramos respectivamente.

   Suponiendo que las poblaciones en estudio son normales con desviaciones típicas conocidas $\sigma_T = 16{,}4$ gramos y $\sigma_{NT} = 7{,}6$ gramos, respectivamente para las ratas tratadas y no tratadas con hormonas,

   a) Obtén un intervalo de confianza al 95 % para la diferencia de ganancias medias de peso.

   b) ¿Cuántas ratas deberían haberse incluido en cada grupo para, con una confianza del 99 %, estimar la diferencia de ganancias medias de peso con un error inferior a 1 gramo? (Supón que los dos grupos son del mismo tamaño)

4. En un estudio sobre las cualidades nutricionales de las *comidas rápidas*, se ha medido la cantidad de grasa en 35 hamburguesas elegidas al azar en una cadena de restaurantes, observándose un contenido medio de 30.2 gramos.

   Supuesto que el contenido de grasa en un hamburguesa es una variable aleatoria con distribución normal de desviación típica conocida $\sigma = 3{,}8$ gramos,

   a) Obtén un intervalo de confianza al 95 % para el contenido medio de grasa en las hamburguesas servidas por esta cadena de restaurantes.

   b) ¿Cuántas hamburguesas deberían haberse elegido en la muestra para, con una confianza del 90 %, estimar el contenido medio de grasa con un error inferior a 0.1 gramos?

5. Un investigador pretende estimar la proporción de hombres que fuman en exceso y que desarrollarán cáncer pulmonar en los próximos años.

   El investigador desea seleccionar cierto número de hombres que hayan fumado mucho durante los últimos años y observarlos en los años venideros para saber cuántos desarrollarán cáncer pulmonar.

   ¿Cuál debe ser el tamaño de la muestra para que, con una probabilidad aproximada de 0.95, la proporción muestral se encuentre a menos de 0.02 unidades de la proporción verdadera?

6. Una encuesta a 100 votantes para conocer sus opiniones respecto a dos candidatos, muestra que 55 apoyan al candidato A y 45 al candidato B. Se pide:

   a) Calcula el intervalo de confianza para proporción de votos de cada candidato, con un nivel de confianza del 95 %.

   b) Calcula el tamaño muestral para que una fracción de 0.55 partidarios de A permita afirmar que será elegido con un nivel de confianza del 99 %.

Problemas de: Estadística descriptiva bidimensional

1. En una encuesta en la que se ha entrevistado a 480 familias, se han obtenido para las mismas los siguientes datos sobre ingresos mensuales (X) y depósitos a la vista en Bancos y Cajas de Ahorro (Y):

Intervalos de Y / Intervalos de X	0-1200	1200-3000	3000-12000	12000-60000
300-600	40	12	8	0
600-900	16	48	12	4
900-1500	8	80	92	20
1500-3000	4	40	72	24

Suponiendo que las marcas de clase son representativas de cada intervalo:

a) Calcula los valores que toman $n_{1\bullet}, n_{2\bullet}, n_{\bullet 2}, n_{\bullet 3}$

b) Expresa, en tanto por uno, los valores que toman: $f_{12}, f_{23}, f_{34}, f_{42}, f_{2\bullet}, f_{3\bullet}, f_{\bullet 2}, f_{\bullet 4}$.

c) Expresa, en porcentaje, los valores que toman: $f_{13}, f_{21}, f_{32}, f_{44}, f_{1\bullet}, f_{3\bullet}, f_{\bullet 4}$.

d) Expresa, en tanto por uno, los valores que toman: $f(X = 450|_{Y=2100})$; $f(X = 750|_{Y=7500})$; $f(Y = 600|_{X=2250})$; $f(Y = 2100|_{X=1200})$

2. Cierta fundación ha concedido 15 becas, distribuidas de la siguiente forma:

Y / X	1	2	3	4
2	1	5	1	2
3	0	2	3	1

X=*número de investigadores* Y=*miles de euros concedidos* .

Calcula:

a) Porcentaje de becas de becas de 3 mil euros y 2 investigadores

b) Porcentaje de becas de 3 mil euros

c) De las becas con 2 investigadores, porcentaje de las de 3 mil euros

3. De una muestra de 29 proyectos de voluntariado de una ONG se recogía información acerca de: X = *número de entidades* ; Y = *número de voluntarios.*

Y / X	1	2	3	4
1	1	2	1	0
3	1	2	3	1
5	1	2	3	6
7	0	1	2	3

Obtén:

a) Distribución de frecuencias absolutas y relativas del número de entidades (X) condicionado a que el número de voluntarios sea superior o igual a tres ($Y \geq 3$)

b) Calcula el número medio de entidades con esa condición.

c) Debido a un cambio en el reparto de proyectos de la ONG, las variables X e Y (entidades y voluntarios) han sufrido las siguientes modificaciones: el número de entidades es ahora $X' = \frac{X+1}{2}$ y por otro lado, el número de voluntarios se ha triplicado: $Y' = 3Y$, ¿Cómo se habrán modificado las varianzas y la covarianza que afectan a estas variables?

4. En una población se tomaron datos de diferentes características, siendo dos de ellas: el número de hijos y la distancia en Km desde el domicilio familiar hasta el lugar de trabajo, para las cuales se obtuvo la siguiente tabla de frecuencias:

D / H	0-2	2-4	4-6
0	24	4	8
1	6	1	2
2	12	2	4

a) Determina el porcentaje de casos en los que una familia tiene 2 hijos y vive a más de 2 Km de su lugar de trabajo

b) Determina el porcentaje de casos en los que una familia tiene 2 hijos, dado que vive a más de 2 Km de su lugar de trabajo

c) Determina la distancia media al lugar de trabajo de las familias con 2 hijos

d) Determina el valor de la mediana para ambas variables.

e) Estudia la dispersión absoluta y la dispersión relativa de las dos variables y comenta los resultados.

f) ¿Es independiente la distancia desde el domicilio al lugar de trabajo del número de hijos?

5. Se ha encuestado a 100 familias en una ciudad, sobre su gasto mensual en ocio (Y), y sus ingresos mensuales (X). Los resultados obtenidos se presentan en la siguiente tabla, donde los ingresos y los gastos están expresados en euros.

Y / X	$[0, 20]$	$(20, 80]$	$(80, 200]$	$(200, 800]$
$[600, 1000]$	4	1	1	0
$(1000, 1500]$	9	8	3	0
$(1500, 2000]$	9	12	20	3
$(2000, 3000]$	6	9	12	3

a) ¿Qué porcentaje de las familias con ingresos no superiores a 1500 euros, gasta en ocio más de 80?.

b) Explica cuál de las dos distribuciones marginales es más homogénea.

c) Razona si X e Y son estadísticamente independientes.

d) ¿Cuál es el ingreso más común?

e) Calcula la covarianza entre ambas variables y explica lo que representa ese valor.

f) Determina si es equitativa la distribución de los ingresos de las familias cuyo gasto en ocio es superior a 80 euros.

g) Obtén los valores que determinan el gasto en ocio del 70 % central de las familias cuyos ingresos no superan los 2000 euros.

Problemas de: Correlación y regresión lineal

1. La siguiente tabla suministra la información sobre la temperatura (en grados Centígrados) en una ciudad, y los litros de cerveza consumidos por los clientes de una cafetería durante 5 días del verano de 2014.

Temperatura	34	28	34	37	40
Litros de cerveza	190	132	200	245	271

   a) Predice la cantidad de cerveza que se consumiría en esta cafetería un día con una temperatura de 35 grados, utilizando un modelo de regresión lineal.

   b) ¿Es bueno el modelo que has construido?

   c) ¿Es fiable la predicción?

2. En un huerto solidario desean encontrar la relación entre la cantidad de agua utilizada para el riego y el rendimiento obtenido en la cosecha. Para ello se seleccionaron siete parcelas con la misma extensión, regándose cada una de ellas con diferentes cantidades de agua.

   La siguiente tabla muestra las cantidades de agua (en metros cúbicos) y los rendimientos (en quintales=cientos de Kg),

Cantidad de agua	120	80	105	70	90	75	105
Rendimiento	133	102	114	89	93	90	123

   a) Predice, utilizando la recta de regresión ajustada a estos datos, el rendimiento que se obtendría con 100 metros cúbicos de agua.

   b) ¿Es fiable la predicción?

   c) Predice, utilizando la recta de regresión ajustada a estos datos, el rendimiento que se obtendría con 200 metros cúbicos de agua.

   d) ¿Es fiable la predicción?

3. En un puesto de un mercado se han observado durante 6 días consecutivos las cantidades de anchoas vendidas en Kilogramos (C) y el precio correspondiente en euros/kg (P). A partir de dichos datos se han elaborado los siguientes resultados:

$$\sum_{i=1}^{6} c_i = 328{,}8 \ ; \ \sum_{i=1}^{6} p_i = 70 \ ; \ \sum_{i=1}^{6} c_i . p_i = 3790{,}515$$

$$\sum_{i=1}^{6} c_i^2 = 19013{,}7213 \ ; \ \sum_{i=1}^{6} p_i^2 = 819{,}3$$

a) Estima los parámetros a y b de la recta de regresión que nos da las cantidades vendidas en función del precio.

b) ¿Es bueno el ajuste?

c) Si el precio de venta es de 12 euros, ¿qué valor estimarías para la cantidad de sardinas vendidas?

4. Para estudiar la relación entre los precios de dos servicios (A y B), se han tomado muestras en 50 establecimientos que los utilizan. Los resultados obtenidos se muestran en la siguiente tabla:

B / A	80	90	96
60	7	8	2
70	3	12	8
80	0	2	8

a) Calcula la covarianza entre los precios de ambos servicios.

b) Estima los parámetros a y b de la recta de regresión que nos da el precio del servicio B en función del precio del servicio A.

c) ¿Es bueno el ajuste?

d) Si el precio del servicio A es de 100 euros, ¿qué precio estimarías para el servicio B?

e) ¿Es fiable esta predicción?

5. Una empresa de servicios a personas inmigrantes toma una muestra de 100 paquetes enviados a través de una determinada empresa de transporte y registra la distancia en miles de Km (D) y el tiempo de entrega en días (T). Realizados unos cálculos los resultados obtenidos son los siguientes:

$$\sum_{i=1}^{100} d_i = 76{,}2 \ ; \ \sum_{i=1}^{100} t_i = 285 \ ; \ \sum_{i=1}^{100} d_i.t_i = 263{,}7$$

$$\sum_{i=1}^{100} d_i^2 = 71{,}043 \ ; \ \sum_{i=1}^{100} t_i^2 = 997{,}5$$

a) Calcula el coeficiente de correlación lineal. ¿Qué significa este valor?

b) Calcula la recta de regresión minimocuadrática que nos da el tiempo de entrega en función de la distancia.

c) ¿Es bueno este ajuste?, ¿por qué?

d) Si se sabe que un paquete se envía a una distancia de 700 Km, ¿qué valor estimarías para su tiempo de entrega?

e) ¿Es fiable esta estimación?

f) Si la distancia (700 Km) aumenta en un 1 %, ¿en qué porcentaje varía el tiempo de entrega?

g) Analiza detalladamente la veracidad de las siguientes afirmaciones:

1) La relación entre D y T es creciente porque el coeficiente de determinación es positivo.
2) El 10 % de la varianza de la variable T no está explicado por la regresión.
3) Si un envío aumenta su distancia en 1000 Km, el tiempo de entrega incrementa en un 3.58 %.
4) Si se cambian las unidades de medida de la variable distancia, la fiabilidad de la regresión no varía.

Problemas de: Análisis estadístico de datos cualitativos

1. En la siguiente tabla figura la clasificación de 10 países, según el número de televisores y de teléfonos por cada 1000 habitantes:

País	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Televisores	4	5	1	10	2	7	3	6	8	9
Teléfonos	1	2	3	10	9	8	7	4	5	6

   Determina si existe asociación entre estos dos indicadores de equipamiento de los países.

2. En una carrera solidaria, 15 atletas (con dorsales numerados del 1 al 15), llegan a la meta en los siguientes puestos:

   7, 5, 6, 9, 10, 12, 1, 8, 2, 3, 11, 14, 4, 15, 13

   Determina el grado de asociación entre el puesto de llegada y el número del dorsal.

3. De una población de 100 personas se ha observado que 30 de ellas están en paro. Los padres de 11 de estas 30 personas también están en paro. Estudia si el paro es una situación que se reproduce dentro de las familias, teniendo en cuenta que de las 100 personas observadas, 40 tienen padres desempleados.

4. Para probar la efectividad de una vacuna se realizó un experimento en un grupo de personas, vacunando a una parte de ellas y observado a las que contrajeron la enfermedad. Los datos se recogen en la siguiente tabla:

	Vacunados	No vacunados
Enfermaron	13	236
No enfermaron	215	450

   Estudia la efectividad de la vacuna:

   a) Estudiando la tabla de contingencia.

   b) Midiendo el grado de asociación.

5. Se ha clasificado a los asistentes a una reunión en función del sexo y de su puntualidad y se obtuvieron los siguientes resultados:

	Puntual	No puntual
Hombre	25	12
Mujer	30	21

   Estudia la asociación o independencia de estas variables:

a) Estudiando la tabla de contingencia.

b) Midiendo el grado de asociación.

6. Se quiere estudiar si existe influencia entre el método de enseñanza (tutorial o presencial) y las calificaciones obtenidas por los estudiantes. Para ello se tomó una muestra de dos grupos de alumnos que realizaron el mismo curso, mediante los dos métodos de enseñanza, y se obtuvieron los siguientes resultados:

	Suspenso	Aprobado	Notable	Sobresaliente
Tutorial	26	25	22	17
Presencial	30	25	37	28

Determina, mediante la prueba de la χ^2, si existe influencia del método de enseñanza en las calificaciones.

7. Los resultados de una muestra tomada en una ciudad sobre el tipo de vehículo y el sexo de sus propietarios es el siguiente:

	Utilitario	Familiar	Berlina	Deportivo	Monovolumen
Hombre	537	85	12	14	6
Mujer	155	11	3	6	1

Determina si podemos aceptar o no la independencia entre los dos atributos.